

1. What Does the SANS Mean?

Levine S¹, Leucht S²

¹University of Haifa, Department of Community Mental Health, Israel, ²Department of Psychiatry and Psychotherapy, Technische Universität München. Klinikum rechts der Isar, Ismaningerstr. 22, 8167, Muenchen, Germany

Methodological Question Being Addressed: What is the meaning of severity and response scores on the SANS?

Introduction: The treatment and measurement of negative symptoms are currently at issue in schizophrenia, but the clinical meaning of symptom severity and change are unclear. Thus the current study aims to offer a clinically meaningful interpretation of severity and change scores on the Scale for the Assessment of Negative Symptoms (SANS).

Methods: Patients were intention-to-treat participants (n=383) in two double-blind randomized placebo-controlled clinical trials that compared amisulpride with placebo for the treatment of predominant negative symptoms. Equipercetile linking was used to examine extrapolation from (a) CGI-S to SANS severity ratings, and (b) CGI-I to SANS percentage change (n =383). Linking was conducted at baseline, 8-14 days, 28-30 days, and 56-60 days of medication.

Results: Across visits, CGI-S ratings of ‘not ill’ linked to SANS scores of 0-13, and ranged to ‘extreme’ ratings that linked to SANS scores of 102-105. The relationship between the CGI-S and the SANS severity scores assumed a linear trend (1= 0-13, 2=15-56, 3= 37-61, 4= 49-66, 5= 63-75, 6= 79-89, 7= 102-105). Similarly the relationship between CGI-I ratings and SANS percentage change followed a linear trend. For instance, CGI-I ratings of ‘Very much improved’ linked to SANS percent change of -90 to -67, ‘much improved’ with -50 to -42, and ‘minimally improved’ with -21 to -13.

Conclusion: The current results contribute to the debate surrounding negative symptoms by providing clinical meaning to SANS severity and change scores and so offer direction regarding clinically meaningful response cut-off scores to guide treatment targets of predominant negative symptoms.

Disclosures: In the last three years Stefan Leucht received speaker/consultancy/advisory board honoraria from SanofiAventis, BMS, Alkermes, EliLilly, Essex Pharma, AstraZeneca, Janssen/Johnson and Johnson, LundbeckInstitute, Medavante and Pfizer. EliLilly provided medication for a clinical trial with Stefan Leucht as the principal Investigator. Stefan Leucht received funding from the German Ministry of Education and Research for a clinical trial and systematic review. The American College of Neuropsychopharmacology has provided a travel grant to attend the ACNP meeting in 2010. Stephen Levine has received research support, and/or consultancy fees and/or travel support from F. Hoffmann-La Roche and Eli Lilly.

2. Latent Profile Analysis of Anxious-Depression among Hispanic/Latinos: Results from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL)

Camacho A¹, Gonzalez P², Buelna C², Emory KT², Talavera GA², Castañeda SF², Espinoza RA², Daviglius ML³, Garcia JE³, Howard AG⁴, Perreira KM⁴, Isasi CR⁵, Roesch S²

¹Departments of Family, Preventive Medicine & Psychiatry, University of California San Diego, ²Institute for Behavioral and Community Health, Graduate School of Public Health, San Diego State University, ³Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, ⁴Collaborative Studies Coordinating Center, University of North Carolina at Chapel Hill, ⁵Department of Epidemiology and Population Health, Albert Einstein College of Medicine

Methodological Question Being Addressed: To determine if there is a common grouping of symptoms from the CES-D and STAI scales that may account for an anxious-depression emotional trait/state in Hispanic/Latinos.

Introduction: Recent findings indicate that anxiety and depression co-occur at high rates, therefore suggesting a new spectrum of anxious-depression emotional states. The upcoming version of the Diagnostic and Statistical Manual for Mental Disorders from the American Psychiatric Association (DSM-5) emphasizes the importance of looking at psychiatric symptoms within a spectrum rather than categorical approach. Studies have reported that some Hispanic/Latino (H/L) patients, present with mixed anxious-depression symptoms. Attempts to subtype anxiety and depression have resulted in minimal success and have revealed mixed classes that do not apply to patients seen in clinical settings or enrolled in clinical trials. The primary aim of this study was to determine if there was a common grouping of symptoms from the CES-D and STAI scales that accounted for a latent subgroup of participants who may endorse high levels of both anxiety and depressive symptoms.

Methods: Data included baseline characteristics of participants from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). Measures included socio-demographic variables that included different Hispanic/Latinos subgroups, depression (Center for Epidemiologic Studies (CES-D)] and, Anxiety (Spielberger Trait Anxiety Inventory [STAI]). Latent profile analyses categorized individuals on the basis of their responses to the CES-D and STAI. It was hypothesized that 3 different classes would emerge: anxious, depressed, and a class that had both anxious and depressive symptoms. We conducted Latent Profile Analyses (LPA) accounting for strata, clustering, and weighting in M-Plus. The indicators included the 10 CESD items and 10 STAI items with responses ranging from 0 (almost never) to 3 (always). LPA with 2 to 6 class structures were performed to determine the best fitting model. Assessment of model fit included the AIC, sBIC, Entropy and the Lo-Mendel-Rubin likelihood ratio test set at a 0.05 threshold.

Results: Out of 16,064 participants, 52% are female with a mean age of 40.2 years. Hispanic subgroups were 42% Mexican, 17% Cuban, 16% Puerto-Rican, 9% Dominican, 7% Central-American and 5% South-American. Thirty seven percent had greater than high school education, 28% had a high school education and 34% had below high school education.

The Lo-Mendel-Rubin likelihood ratio was significant up to 5 classes. Although our hypotheses that three subgroups representing elevated depression, anxiety and mixed anxiety/depression scores would emerge from the LPA was not supported, a 3-Class model did result in a better fit based on our proposed hypothesis of an anxious-depression spectrum and it also was the most interpretable class structure. The 3-Class model fit statistics were: AIC=787915.541, sBIC= 788285.066, Entropy=0.898, and LRT<0.05. The 3 classes were: 1) a low anxiety and depression group, 2) a moderate anxiety and depression group and, 3) a high anxiety and depression group, suggesting a spectrum of anxious-depression symptomatology.

Conclusions: The results suggest that participants clustered into either a low, moderate or high class of anxious-depression symptoms providing a good fit for the data. Future studies should evaluate the clinical implications of these latent profiles representing varying levels of mixed anxiety and depression symptoms.

Disclosure: Funded by T32 HL079891-06 UCSD Integrated Cardiovascular Epidemiology Fellowship and The Hispanic Community Health Study/Study of Latinos, which was supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01-HC65235), Northwestern University (N01-HC65236), and San Diego State University (N01-HC65237). The following Institutes/Centers/Offices contribute to the HCHS/SOL through a transfer of funds to the NHLBI: National Center on Minority Health and Health Disparities, the National Institute of Deafness and Other Communications Disorders, the National Institute of Dental and Craniofacial Research, the National Institute of Diabetes and Digestive and Kidney Diseases, the National Institute of Neurological Disorders and Stroke, and the Office of Dietary Supplements. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

3. ASPECT-R: A Tool to Rate the Pragmatic (Effectiveness) / Explanatory (Efficacy) Characteristics of a Clinical Trial Design

Bossie C¹, Alphas L¹

¹Janssen Scientific Affairs, LLC

Methodological Question Being Addressed: Development of an instrument to assess the pragmatic (effectiveness) / explanatory (efficacy) characteristics of clinical studies.

Introduction: Efficacy or explanatory trials ask whether an intervention works. To minimize confounds, these trials are almost always done under highly controlled and defined conditions. Effectiveness or pragmatic trials seek to answer whether an intervention works under usual or real-world conditions. Both approaches have value, and many trials have aspects of both designs. ASPECT-R (A Study Pragmatic-Explanatory Characterization Tool - Rating) is an instrument derived from an existing tool that was developed to assist researchers in designing trials that are either more pragmatic or more explanatory (PRECIS, Thorpe et al. 2009, Tosh et al. 2011). The feasibility and limitations of using ASPECT-R to retrospectively rate an interventional trial in schizophrenia is described.

Methods: ASPECT-R consists of 8 design domains deemed important in characterizing the pragmatic vs explanatory characteristics of a study. These domains are related to participant eligibility, intervention(s) flexibility,

practitioner expertise, outcomes, follow-up intensity/duration, and participant compliance. This tool differs from the early modifications of PRECIS in that descriptive anchors were developed for the rating of each domain (0=extremely explanatory to 6=extremely pragmatic; total study score range: 0-48). A study anticipated to have both pragmatic and explanatory characteristics (CATIE, Lieberman et al. 2005) was rated by the authors by consensus. Authors discussed the rationales for ratings, identified areas of dispute, and identified scale limitations. Results will be presented descriptively.

Results: The total ASPECT-R score for this study was 22. Domain ratings illustrated that this study had aspects of both designs: 2 domains were rated as more pragmatic than explanatory (participant eligibility criteria and primary trial outcomes); 3 domains had an intermediate rating (flexibility for the experimental and comparison interventions and participant compliance); and 3 domains were rated as more explanatory (practitioner expertise for the experimental and comparator interventions, and follow-up intensity/duration). A limitation to the application of the tool was that relevant information was sometimes poorly documented or unavailable for some domains.

Limitations include that the tool is still under development and inter-rater reliability and other validation work is incomplete. In addition, some domains may be somewhat overlapping, sharing non-unique contributions. Although ASPECT-R does not consider the quality of the study conduct, design, or interpretation relative to the objective, a companion tool is being developed to address this issue.

Conclusion: This work demonstrated the feasibility of using ASPECT-R to rank a study's key design domains on the pragmatic:explanatory continuum. Limitations, including validation with independent raters, will be addressed in follow-up work.

Supported by Janssen Medical Affairs LLC

4. Assessment of the Bipolar Inventory of Symptoms Scale in a Sample of Older Persons with Bipolar Disorder

Bowden C¹, Arnold J¹, Vivek Singh¹, Burns M¹, Tohen M¹

¹University of Texas Health Science Center at San Antonio

Methodological Question Being Addressed: Does the BISS perform as well as other rating scales in assessing symptomatology in older adults with bipolar disorder?

Introduction: The Bipolar Inventory of Symptoms Scale (BISS) is a comprehensive semi-structured rating scale that has demonstrated good reliability and validity. No studies have been conducted assessing its reliability among an older population. This study aims to examine the reliability and validity of the BISS among an older sample with bipolar disorder (BD). We hypothesized that the older bipolar group would have higher ratings on impulsive behavior, risky behavior and affective lability.

Methods: We conducted the BISS with an older population of subjects ≥ 60 in age, including 25 subjects with bipolar I or II disorder, any mood state, and 25 healthy controls matched on age, ethnicity and gender. Older subjects with BD were compared to a younger group of subjects (18 -59 years, n=93) with BD. We carried out one way ANOVAs using log-transformed data for homogeneity of variance. We used Fisher's Least Significant Difference for pairwise multiple comparisons of group means.

Results: Older subjects were more likely to be female (76% versus 51%). There were no differences between the older and younger group of subjects in BD type, 73% vs 77% respectively. Healthy controls and the younger group of BD subjects were more likely to be married.

The older bipolar sample had very similar results on the reliability of items and scales as the younger subjects: Chronbach's alpha, for the YMRS = .80, the BISS Manic Scale = .79, and BISS Mania Factor .82, MADRS = .85, BISS Depression Scale .93 and BISS Dep Factor = .91.

The older group of BD subjects had significantly higher scores on all rating scales and subscales (YMRS, MADRS, BISS Total, Depression and Mania Scales and 5 factors) compared to the healthy controls except for the psychosis factor. In the older BD group, 14 were euthymic while 11 subjects met the DSV-IV defined criteria for a syndromal mood state. For the younger group, 26 were euthymic and 72 met the criteria for a syndromal mood state. The older euthymic group had a significantly higher average GAF score of 84 while the younger euthymic group had a GAF score of 72. The older subjects in a syndromal mood state had an average GAF score of 71 while the average GAF

score for the younger subjects in a syndromal mood state was 56. The euthymic younger group scored higher on the BISS depression subscale ($p=.047$) and the BISS irritability factor (.004) than the euthymic older group. Among subjects in a syndromal mood state, the younger group had systematically worse scores across measures. The younger group of BD subjects had higher scores on impulsivity, risky and affective lability items, not consistent with our hypotheses.

5. The Development of the Restorative Sleep Questionnaire and Levels of Non-Restorative Sleep Reported by ISCTM Conference Attendees

Macek T1, Morlock R2, Roth T2

¹Takeda Development Center Americas, Inc., Deerfield IL ²Henry Ford Hospital, Detroit MI

Methodological Question Being Addressed: The development of a measure for restorative and non-restorative sleep and its performance in a population of scientific meeting attendees, compared to the general population

Introduction: Insomnia symptoms are categorized as difficulty initiating sleep (DIS), difficulty maintaining sleep (DMS) and non-restorative sleep (NRS). Patient report and objective measures exist for DIS and DMS; however, limited options exist to assess NRS. A session at the Autumn 2011 ISCTM Conference addressed the methodological issues in the measurement of NRS. The session consisted of an overview of the differential diagnosis, phenomenology, epidemiology, and characteristics of NRS in the general population, the methodological challenges in conducting clinical studies for NRS treatment, European and US regulatory perspectives on NRS, as well as the development of measures of NRS, namely the Restorative Sleep Questionnaire (RSQ). The RSQ was developed to assess perceptions of the restorative aspects of sleep and the consequences of non-restorative sleep. The RSQ was developed using patient and clinician focus groups and the psychometric properties were assessed in a number of studies including objective polysomnographic (PSG) assessments. To assess levels of NRS within a population of ISCTM meeting attendees, prior to the meeting, all registered attendees were asked to complete the RSQ. Here we describe the development and psychometric properties of the RSQ and levels of NRS within ISCTM respondents of the Autumn 2011 conference.

Methods: Development of the RSQ started with expert/clinician interviews followed by focus groups consisting of good and poor sleepers who identified key concepts relevant to the restorative aspects of sleep both on awakening and its consequences throughout the subsequent day. Cognitive debriefing tested the understandability of resulting questions and response categories to subjects with and without complaints of NRS. Psychometric evaluation of the RSQ was carried out using results from a telephone survey of the general population and clinical trials using polysomnography in patients with insomnia and in normal sleepers. Results across patients groups were compared (patients with insomnia vs. normal sleepers). For ISCTM meeting respondents, informed consent was obtained and data collected via a web-based questionnaire between September 14 and 28, 2011.

Results: Item-total correlations (corrected for overlap) for the RSQ Ranged from 0.40 to 0.84. Coefficient alphas were consistently above 0.90 across studies. More restorative sleep was associated with lower sleepiness scores ($r's = -0.28$) and better SF-36 PCS ($r's = 0.28$) and MCS ($r's = 0.42$) scores. The RSQ was able to differentiate between PSG confirmed normal sleepers and patients with DIS, DMS and DIS as well as DMS. The RSQ was sent to 148 individuals registered for the 2011 NRS ISCTM Conference. The average age of ISCTM respondents was 50.34 years and 61% were male. In general, ISCTM respondents reported themselves to be highly educated and in good health with minimal sleep complaints. Total sleep time averaged 6.8 hours for ISCTM respondents. 49% of ISCTM respondents reported that they believe NRS is a distinct entity and 45% were unsure. Total RSQ scores were consistent between ISCTM respondents and previous studies. ISCTM respondents with self-reported DIS or DMS had RSQ scores similar to those reported in the validation studies.

Conclusions: The RSQ is a rigorously developed and validated instrument for use in clinical trials. The tool has acceptable psychometric properties and is able to differentiate between objectively measured sleep complaints of DIS and DMS. RSQ scores in ISCTM meeting respondents with or without sleep complaints were similar to data obtained previous studies.

6. Improvement in Ratings Behavior Over Time in Four Separate Placebo Controlled MDD Trials

Busner J¹, Randall P², Wilson E³, Tummala R³

¹ Bracket, and Penn State College of Medicine, Department of Psychiatry, ² Quintiles, and University of California, San Diego, Department of Psychiatry, ³ AstraZeneca

Methodological Question Being Addressed: Can rater errors improve over time in clinical trials and is such improvement dependent upon visit type (baseline vs. non-baseline)?

Introduction: Recent work has shown baseline to later visit improvements in internal consistency (Cronbach's alpha) of efficacy measures in MDD ^{1,2}, schizophrenia ³, and bipolar disorder ⁴ clinical trials that included external rater oversight and remediation interventions. This is consistent with work showing decreased rater scale usage errors in initial (i.e., screening or baseline) compared with final subject visits ⁵. It is unclear whether and to what degree such improvements are due to factors unique to problematic baseline/screening visits as has been suggested ⁶ or due to time independent of particular visit. We conducted a retrospective analysis of rater errors pre and post chronologic midpoints for four separate multinational clinical trials. The methodology allowed us to examine errors as a function of time rather than visit type.

Methods: Four large separate industry sponsored double-blind placebo-controlled clinical trials were examined retrospectively. Each trial had made use of daily computerized clinical data quality checks for potential rater scale misuse, with external oversight and clinical remediation of site raters provided throughout the trials as errors occurred.

Results: The proportion of baseline visits in the first chronological halves of the four studies was 48%, 45%, 50%, and 56%, respectively, ensuring for at least three of the studies that a preponderance of Baseline visits was not contained in the first half of the study. For each of the four studies, flag rates decreased significantly in the second half of the trial compared with the first half of the trial [χ^2 's (1)=96.5, 225.3, 1000.7, 762.9, respectively, all p's < .0001].

Conclusions: In four individual industry sponsored trials, clinical data errors lessened significantly as a function of time in surveillance program. As baseline visits were well distributed across time periods, the findings suggest that time in program improves performance, independent of visit type. The findings lend support to the provision of ongoing rater oversight and remediation throughout the course of large multinational trials.

Disclosures: Dr. Busner is a full time employee of Bracket; Dr. Randall is a full time employee of Quintiles; Mr. Wilson and Dr. Tummala are full time employees of AstraZeneca.

References:

- 1) Busner, J. Daniel, D.G., Montgomery, S.A., Bartko, J.J. (2012). Lower Cronbach's Alpha at Baseline Than Next Visit in MDD Studies With and Without Separate Inclusionary Scales. Presented at the NCDEU Annual Meeting, May 29-June 1, Phoenix, AZ.
- 2) Busner, J., McNamara, C., Oakley, M., Platco, K., Montgomery, S. (2011). Surveillance strategies for enhancing data quality in adjunctive psychopharmacotherapy trials. Presented at the American Psychiatric Association Annual Meeting, May 14-18, Honolulu, HI.
- 3) Daniel, D., Busner, J., McNamara, C. (2011). Internal Consistency of Ratings Improve and Error Rates Decrease with Ongoing Monitoring and Feedback in an International Schizophrenia Clinical Trial. Presented at the 13th International Congress on Schizophrenia Research, April 2-6, Colorado Springs, CO.
- 4) Sachs GS, Edman S, Debonis D. (2011) Variation in MADRS Reliability Over the Course of Randomized Controlled Trials: Comparison of Computer vs. Site-based rater administration. Poster presentation, 51st New Clinical Drug Evaluation Unit meeting, Boca Raton, FL.
- 5) Busner, J., Vieta, E., Bartko, J., Bealer, R., Spear, C. Global clinical trials: quality assurance methodology for site based, global clinical trial ratings. (2008). *The International Journal of Neuropsychopharmacology*, 11(Supplement 1):311.
- 6) Dunn J. Comparison of Site-based versus Centralized Ratings in a Clinical Trial of Generalized Anxiety Disorder. (2010). 163rd annual meeting of the American Psychiatric Association, New Orleans, May 24.

7. Are Intra Class Correlations (ICCs) Enough to Assess Rater Reliability? No!

Yavorsky C^{1,2}, Opler M^{1,3}, DiClemente G², Khan A^{1,4}, Rothman B¹, Lucic L^{1,5}, Jovic S¹

¹ProPhase, LLC, ²CROnos CCS, ³New York University School of Medicine, ⁴Nathan S. Kline Institute for Psychiatric Research, ⁵Pratt Institute

Methodological Question Being Addressed: As psychiatric rating scales comprise of complex items and sometimes subjective judgments, conventional methods to assess reliability needs expansion. What are the components of reliability that should be examined to ensure that clinical trials are understanding and building on the limitation of traditional reliability and consistency findings?

Background: When assessing interrater reliability, many rely only on satisfactory ICCs (≥ 0.75) as the criteria of a skilled rater. Because psychiatric assessments often consist of complex items and may involve subjective judgments, traditional approaches to reliability are insufficient.

Methods: Raters scored the same interview for one of three scales as part of rater certification criteria for clinical trials –Positive and Negative Syndrome Scale (PANSS), Montgomery-Asberg Depression Rating Scale (MADRS), and Young Mania Rating Scale (YMRS) (n = 100, each). Methods of quantifying inter-rater reliability were classified into three categories: 1) ICCs; 2) percent agreement; 3) agreement with core symptoms (items) of the disorder.

Results: PANSS: ICCs with Gold Standard scores ranged from 0.129 to 0.901. 74.00% raters achieved an ICC ≥ 0.75 . When adding any one of the criteria listed above, reliability declined from 59.00% - 67.00%. MADRS: ICCs = 0.324 to 0.900. 78.00% raters achieved an ICC ≥ 0.75 . When adding any one of the criteria listed above, reliability declined from 54.00% - 68.00%. YMRS: ICCs = 0.199 to 0.892. 72.00% raters achieved an ICC ≥ 0.75 . When adding any one of the other criteria listed above, reliability declined from 52.00% - 65.00%.

Conclusions: Although rater reliability limits the value of the scores derived from a scale, it is only one aspect of the broader question of consistency of scores. An ICC is a necessary but not sufficient condition for score reliability in psychiatric scales, because high interrater reliability does not imply that overall scale reliability is satisfactory.

Keywords: Reliability, PANSS, MADRS, YMRS

Disclosures: No authors have conflicts of interest related to this study to disclose. Yavorsky, C and DiClemente G are full time employees of CROnos CCS, Opler M is the owner of ProPhase LLC; Khan A is employed part time at Manhattan Psychiatric Center and at ProPhase LLC, and volunteers at Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY. Rothman B and Jovic J are full time employees at ProPhase LLC. Lucic L is a full time employee at ProPhase LLC and a visiting professor at Pratt Institute. NY, NY.

8. Validation and Variable Strength Testing of a Subject Enrollment Monitoring System

Brown, BR¹, Yavorsky, WC², DiClemente, G^{1, 2}, Smith, AD¹

¹Kashima Consulting Associates, ²CROnos Clinical Consulting

Methodological Question Being Addressed: Can algorithms be used to detect professional, recycled and inappropriate subjects? Are there particular subject characteristics that lend greater sensitivity and specificity to the algorithmic results?

Introduction: Despite acknowledgement and concern in the scientific community, to date there has been little done to help detect professional or recycled subjects participating in clinical trials. An innovative Professional, Recycled and Inappropriate Subject Monitoring system (PRISM) designed to analyze patient eligibility with the goal of significantly mitigating the risk of including these patients was validated. Two primary hypotheses were tested: 1) the system shall identify patterns amongst subjects with obvious or non-obvious associated characteristic sets, and 2) the system should not identify patterns amongst subjects with dissimilar sets of characteristics. Also of interest was exploring the power that specific exploratory subject characteristics may contribute to value of these eligibility searches.

Methods: A population of 10,000 CNS clinical trial subjects with characteristics typically collected over the course of a clinical trial was selected and stored within our database. One thousand additional subject characteristic sets (“test subjects”) were created. Five hundred test subjects were “true positives,” meaning characteristics were associated, though not identical, to a subject within the population. These test subjects were created to resemble professional or recycled subjects, changing variables as they may vary in the real world with time and deception. The remaining 500 test subjects were “true negatives,” meaning characteristics were dissimilar in one or more characteristics to the subjects in the population. A proprietary similarity search algorithm was used to find and rank order similar subjects across a range of characteristics.

Results: Results show very high detection rates of inappropriate subjects with PRISM. Detection of “true positives” varies with the number of subject characteristic variables used in the searches (94%-99.4%). False negatives and false positives are present within all searches (0.6-4.5%), even with the use of deception-resistant variables like blood type, however all conditions seem to keep false notifications at an acceptable level. Exploratory subject characteristic variables all brought improvement to the basic variable detection results across all categories (true positive, false positive and false negative). Blood type and wrist circumference were the variables that resulted in the largest improvements in detection rates.

Conclusions: PRISM appropriately identified the professional, recycled or inappropriate subjects, thus use of this system may help to detect such subjects in a clinical trial prior to randomization and thereby eliminate the risk that may come along with them. The resulting enhancement in the quality of enrolled subjects will lead to more efficient trials of CNS treatments with increased validity and power in results. Sponsors may wish to consider the collection of wrist circumference or blood type subject characteristics in order to have optimal performance of the monitoring system.

9. Feasibility of Centralized Ratings for Mental Health Safety Screening in a Dermatology Trial

Williams JBW^{1,2}, Davis D¹, Popp D¹, Gross JA³, Salvucci D¹, Detke MJ^{1,4}

¹MedAvante, Inc., ²Department of Psychiatry, Columbia University, ³Cipher Pharmaceuticals, ⁴Department of Psychiatry, Indiana University

Methodological Question Being Addressed: Clinical trials in non-CNS indications frequently include assessments of psychopathology and safety signals. Non-mental health sites may not be equipped to diagnose exclusionary mental disorders or follow-up on suicidality. The purpose of this study was to evaluate the feasibility of centralized remote expert clinicians assessing mental disorders and suicidality in subjects in a dermatology trial.

Introduction: Centralized ratings by videoconferencing or telephone have proven feasible for assessment of adult psychiatric diagnosis, symptom severity, and suicidality, and may be used for safety assessments in non-psychiatric trials with sites that do not employ staff experienced in psychiatric assessment (e.g., dermatology). Centralizing assessments with mental health experts enables immediate clinical follow-up and actionable diagnostic support for investigators. This study examines the feasibility and acceptability of using centralized ratings in a Phase III dermatology clinical trial as a means of evaluating treatment effects and establishing safety indicators in adults and adolescents.

Methods: 7988 assessments were performed via telephone on 1127 subjects who were patients at dermatology offices and enrolled in this clinical trial of a medication for their dermatologic condition. These assessments were done initially to assess study eligibility, within study to determine treatment effects, and post-study to assess treatment sequelae. Subjects were adults (n=630) and adolescents (n=497). At screening, centralized raters administered the SCID-CT, C-SSRS Lifetime and Last Year to assess suicidality, and PHQ-8 for depressive symptom severity. At monthly visits, central raters performed the C-SSRS Emergent, PHQ-8, GAD-7 and items designed to detect emergent psychotic symptoms.

Results: *Screening:* 34 subjects (3%) were excluded on the basis of SCID-CT diagnosis. Of these, 27 (2.4%) were excluded for a major depressive episode and one for hypomania in the past year, one for a lifetime major depressive episode, and five for a lifetime psychotic episode. Based on diagnosis or severity, subjects could be classified as being in no need of mental health services, or having mild psychiatric symptoms (referred to local mental health service provider; n=33), moderate (immediate referral for psychiatric evaluation; n=17), or severe (immediate escort to emergency room; n=0). At screening one subject reported suicidal ideation on the C-SSRS, 1% reported self-injurious behavior (n=10), and 0.5% reported suicidal behavior in the last year (n=5). Scores on the PHQ-8 at screening ranged from 0–21 (M=1.02; SD=1.89). 54% of subjects scored a 0 on the PHQ-8 (n=612) and eight subjects had scores greater than 10.

Follow-Up: No subjects reported suicidal ideation or behavior at any of the 6861 follow-up assessments. One subject reported self-injurious behavior and two reported emergent psychotic symptoms. PHQ-8 and GAD-7 scores were stable within each subject over the course of the study.

Conclusions: This study established the feasibility and acceptability of routine screening and monitoring of psychopathology and suicidality by central raters in a non-psychiatric population. Central raters identified subjects

who did not qualify for entry at the beginning of the study because they had active suicidal ideation and significant active or recent mood disorder in the last year, or a lifetime incidence of psychosis. These subjects were excluded from the study and referred for clinical care. Throughout the study, central raters identified cases of emergent psychosis and mood symptoms.

Disclosures: Drs. Williams, Popp, Davis, and Ms. Salvucci are full-time employees of MedAvante, Inc. Dr. Gross is Employed as Vice President of Scientific Affairs for Cipher Pharmaceuticals. Dr. Detke reports being an employee and major stockholder of MedAvante, Inc. and has provided expert testimony for Eli Lilly, and received fees for consultation and participation on advisory boards for NIH, Roche, Sonkei, Phine Pharmaceuticals, Columbia NW Pharmaceuticals, Insight Neuropharma, Inc., and Jeevan Scientific, Inc

10. Integration and Impact of External Confirmation of Subject Eligibility in an Acute Schizophrenia Proof of Concept Study

DeMartinis, N.¹, Boyer, S.¹, Banerjee, A.¹, Kumar, V.¹, Ereshefsky, L.², Targum, S.³

¹Pfizer Inc., Neuroscience Research Unit, ²Parexel International, ³Clintara

Methodological Question Being Addressed: Minimizing placebo response and improving efficacy signal detection.

Introduction: A Phase 2A study was conducted to evaluate efficacy and safety of PF-02545920 in the treatment of acute exacerbation of schizophrenia. Efforts to minimize placebo response and increase signal-detection focused on a two-stage ‘gatekeeping’ strategy to maximize enrollment of an appropriate patient population. A methodology for external, site-independent confirmation of subject eligibility was developed to operationalize this goal. Before randomization, screening data was evaluated by an external reviewer to ensure that all symptom and course of illness entry criteria were met. This approach aimed to optimize quality of subjects entering the study by leveraging research site strengths, maintaining engagement of investigators, and minimizing operational burden at the site and for the sponsor.

Methods: The study was a double-blind, parallel-group, placebo and positive-controlled, randomized, multi-center, trial investigating 5 mg and 15 mg (titrated) doses of PF-02545920 BID, 3 mg (titrated) of risperidone BID, and placebo. The primary efficacy endpoint was change from baseline in PANSS total score at week 4. The Clinical Validation Inventory for Study Admission (C-VISA) was a worksheet developed to summarize all clinical entry criteria and supporting evidence, incorporating information from multiple sources. The external review process encompassed the following: 1) key symptom and diagnostic interviews conducted by site-based raters were graphically digitized and audio recorded via smart-pen, and the investigator completed the C-VISA worksheet; all were uploaded to a central site, 2) an external reviewer examined the uploaded files, and 3) provided documented review within 72 hours. All non-approvals were escalated to a Tier 2 expert reviewer for the final eligibility decision, which included direct contact with the investigator (adjudication).

Results: The mean time for external review was 53 hours from receipt of the electronic submission to final determination. The addition of external review did not adversely affect enrollment; 19.4% of the Tier 1 reviews were escalated to Tier 2. 18 of 49 Tier 2 reviews were adjudicated as screen failures, for an overall screen failure rate of 7.1%. Reasons included co-morbid conditions, confounding factors, and insufficient rater interview/unreliable subject. 31 Tier 2 reviews were approved for randomization subsequent to adjudication with the site investigator. 74% of randomized Tier 2 approvals completed the study, comparable to the 76% completion rate for all subjects. The study results for the Tier 2 approvals were similar to those for the Tier 1 approvals.. The active-control demonstrated efficacy comparable to recent literature findings, and the placebo response and primary endpoint variability compared favorably to observations in similar recent studies demonstrating assay sensitivity.

Conclusions: Use of external confirmation of subject eligibility demonstrated exclusion of inappropriate subjects that might have otherwise been randomized, with no decrease in study operational efficiency. The study completion rate, endpoint variability, and placebo response indicated enrollment of a high quality study population. Utilization of this external review procedure demonstrated that site-independent confirmation of subject eligibility based on site-generated data can positively impact study integrity and data quality in a time-efficient manner while supporting investigator engagement.

11. Clinically Relevant HAM-D and YMRS Scores in the Schizoaffective Population

Turkoz I¹, Fu D-J², Bossie CA², Alphas L²

¹Janssen Research and Development, LLC, Titusville, NJ, USA, ²Janssen Scientific Affairs, LLC, Titusville NJ, USA

Methodological Question Being Addressed: Identify cut-points on continuous scales that measure mania and depression that map to corresponding manic and depressive subscale scores for the Clinical Global Impressions–Severity for Schizoaffective Disorder (CGI-S-SCA) scale.

Introduction: Clinical trials of schizoaffective disorder (SCA) require assessment of psychosis, depression, and mania symptoms with an array of scales including the PANSS, YMRS, and HAM-D-17. However, clinically relevant ranges (cut-points) of YMRS and HAM-D-17 are not well established in the SCA population. This analysis explored the relationship between ratings on YMRS and HAM-D-17 total scores and those on manic and depressive subscales of the CGI-S-SCA as CGI scores may more readily reflect a clinically meaningful state to clinicians.

Methods: Data from two 6-week, randomized, placebo-controlled studies of oral antipsychotic versus placebo in symptomatic subjects with SCA (N=627) were used to identify cut-points of the YMRS and HAM-D-17 by exploring their relationships to the ratings of the CGI-S-SCA manic and depressive domain, respectively. Ranges on scales were generated using 2 methods: a nonparametric approach using discriminant analysis and a parametric approach using generalized linear models. Nonparametric methods require no or very limited assumptions to be made about the data. Sensitivity and specificity analyses were performed to compare the 2 approaches on classification of the mood scales.

Results: Comparison of YMRS scores with the categorical rating on the CGI-S-SCA manic domain using the nonparametric approach demonstrated a sensitivity range of 42.5%-53.9% and a specificity range of 83.2%-95.8%. With this analytic approach, YMRS scores associated with a CGI-S-SCA manic domain=1 (not ill) were 0-6; for CGI=2 (borderline), 7-12; for CGI=3 (mild), 13-19; for CGI=4 (moderate), 20-26; for CGI=5 (marked), 27-37; and for CGI=6/7 (severe/extremely severe), 38-60. The parametric approach demonstrated a sensitivity range of 3.4%-75.8% and a specificity range of 80.9%-99.3%. Using this approach, YMRS scores associated with CGI-S-SCA manic domain=1 were 0-11; with CGI=2, 12; with CGI=3, 13-19; with CGI=4, 20-32; with CGI=5, 33-43; and with CGI=6/7, 44-60.

Comparison of HAM-D-17 scores with categorical ratings on the CGI-S-SCA depression domain using the nonparametric approach had a sensitivity range of 43.9%-66.8% and a specificity range of 76.8%-97.6%. Using this approach, HAM-D-17 scores associated with CGI-S-SCA depression domain=1/2 (not ill/borderline) were 0-7; with CGI=3 (mild), 8-14; with CGI=4 (moderate), 15-20; with CGI=5 (marked), 21-28; and with CGI=6/7 (severe/extremely severe), 29-52. The parametric approach demonstrated a sensitivity range of 8.4%-89.2% and a specificity range of 76.7%-99.9%. Using this approach, HAM-D-17 scores associated with CGI-S-SCA depression domain=1/2 were 0-11; with CGI=3, 12-15; with CGI=4, 16-25; with CGI=5, 26-34; and with CGI=6/7, 35-52.

Conclusion: These analyses identified relationships between ratings on the YMRS and HAM-D-17 total scores and those on the manic and depressive subscales of the CGI-S. The nonparametric approach provided clinically relevant ranges similar to those previously established in the bipolar or depression populations. (Rush et al, *Am J Psychiatry*.2006;163:1905–1917; Furukawa et al, *J Clin Psychopharm*.2007;27:332–333).

Disclosures: LA, IT, CB and DJF are employees of Janssen and Johnson & Johnson stockholders.

Supported by Janssen Scientific Affairs, LLC

12. Circadian Rhythms and Cognition: Using Time of Day Effects To Enhance Signal Detection in Clinical Trials of Pro-Cognitive Therapies for Schizophrenia

Michael Hufford¹, Vicki Davis¹, Nancy Dgetluck², Maria Gawryl², Dana Hilt², Alon Rappaport³, and Richard Keefe^{1,4}

¹NeuroCog Trials; ²EnVivo Pharmaceuticals; ³BiolineRx; ⁴Duke University

Methodological Question Being Addressed: Can taking into account diurnal variation in cognitive functioning help to enhance signal detection in clinical trials of pro-cognitive agents for schizophrenia?

Introduction: Circadian rhythms exert changes in cognitive functioning over the course of the day. Patients with

schizophrenia are known to have profoundly disturbed circadian rhythms that can affect their cognitive functioning. Diurnal variations in cognitive functioning were analyzed by examining the impact of time of day on baseline composite MATRICS Consensus Cognitive Battery (MCCB) scores. Next, post hoc exploratory analyses were conducted in two Phase 2 clinical trials to examine whether taking into account consistency in the timing of neurocognitive administrations between the baseline and endpoint visits could affect signal detection.

Methods: For the diurnal variation analyses, 1,971 baseline MCCB assessments were aggregated across 8 separate schizophrenia clinical trials. The assessments were divided into 2-hour time intervals based on the start-time of the assessments (varying from 8am and 5pm) and then analyzed for differences by time interval. Next, two Phase 2 schizophrenia clinical trials were used to explore the impact of diurnal variation on pro-cognitive signal detection. We separated subjects into those with consistent (± 1 hr) versus inconsistent (>1 hr) timing of neurocognitive battery administrations between their baseline and endpoint visits, and then compared the subgroups.

Results: Time of day exerted a significant effect on baseline composite MCCB scores ($p=.0009$), with composite scores varying more than 7 points over the course of the day. Follow-up contrasts with Bonferroni correction for multiple comparisons revealed significant differences among multiple temporal epochs. Next, analyses examined whether taking into account consistency in the time of day of neurocognitive administrations between the baseline and endpoint visit could affect signal detection. The first clinical trial was a 12-week placebo-controlled trial of an add-on therapy for cognition using the MCCB among subjects with schizophrenia stabilized on antipsychotic therapy recruited from the US ($n=170$). These analyses revealed that the treatment effect continued to favor drug over placebo for subjects with consistent assessment timing, whereas no trend was evident among subjects with inconsistent timing. The second clinical trial was of a broad-spectrum antipsychotic used to treat acutely ill subjects for 6-weeks among patients with schizophrenia recruited from the US, India, and Romania ($n=363$). Cognition was assessed using the Brief Assessment of Cognition in Schizophrenia (BACS). These analyses revealed that the consistent timing group showed a more robust treatment response as compared to the inconsistent timing group.

Therapy (Battery)	Grouping	LSM Change (SE) from Baseline				p-Value for High-Dose vs. PBO
		PBO	Low-Dose	High-Dose	Active-Control	
Add-On (MCCB)	Consistent-timing (n=58)	1.3 (1.5)	4.1 (1.6)	5.9 (1.8)	n/a	.044
	Inconsistent-timing (n=81)	2.8 (1.2)	2.3 (1.1)	2.4 (.97)	n/a	.805
Broad-Spectrum (BACS)	Consistent-timing (n=84)	5.5 (2.5)	7.2 (2.2)	16.4 (2.9)	9.9 (2.3)	.005
	Inconsistent-timing (n=134)	5.8 (1.6)	7.5 (1.7)	9.2 (1.7)	7.1 (1.9)	.394

Conclusions: Cognitive functioning ebbs and flows over the course of the day. Maintaining consistency in the time of day of neurocognitive administrations between visits can help to enhance signal detection in clinical trials of pro-cognitive therapies.

13. Measuring Transitions of Care of Schizophrenia Subjects Released From Incarceration Using Resource Use Questionnaire—Baseline Data

Mao L¹, Benson C², Fastenau J², Ferziger R², Alphs L²

¹Janssen Research & Development, LLC, Titusville, NJ, USA, ²Janssen Scientific Affairs, LLC, Titusville, NJ, USA

Methodological Question Being Addressed: Outpatient services for mentally ill offenders are available but often underutilized. Early and appropriate interventions, including delivery of outpatient services and concurrent effective pharmacologic interventions, may prove beneficial in reducing arrest among adults with serious mental illness. Resource utilization questionnaire (RUQ), which measures outpatient utilization and transitions of care longitudinally, can provide meaningful information in making early and appropriate decisions for intervention.

Introduction: To describe baseline resource utilization by subjects enrolled in an ongoing pragmatic clinical trial. Paliperidone Research in Demonstrating Effectiveness (PRIDE) is a 15-month randomized, active-controlled,

open-label study of paliperidone palmitate compared with oral antipsychotic treatment in previously incarcerated adults with schizophrenia. The primary endpoint is time to a protocol-defined treatment failure, a composite endpoint that includes recidivation to jail or prison, psychiatric hospitalization, increased psychiatric services to prevent imminent psychiatric hospitalization, suicide, and discontinuation of antipsychotic treatment supplementation due to inadequate efficacy and of medication due to inadequate efficacy, safety, or tolerability.

Methods: Subjects randomized by 15 March 2012 were included in this analysis. Baseline resource utilization was assessed using the resource utilization questionnaire (RUQ), completed by study staff through subject interviews and supplemented by available health care, criminal justice, and social services records. Baseline information collected by the RUQ includes sociodemographic classification, outpatient and inpatient services, emergency room (ER) visits, use of emergency medical services (EMS) and contacts with the criminal justice system before study entry during the 12 months before last incarceration and from last incarceration to baseline assessment. Data were analyzed using descriptive statistics.

Results: Mean (SD) age of the 340 randomized subjects was 37.8 (10.51) years; mean (SD) age of first psychiatric diagnosis was 19.9 (7.40) years; 87.6% were male; 18.2% had concurrent diagnosis of substance abuse; 4.8% reported mild and 1.8% moderate suicidal ideation. 44.0% of subjects arrested or incarcerated at least twice in the past 12 months. Arrest was mostly for nonviolent offenses: probation/parole violation (18.0%), drug charges (16.5%), disorderly conduct/vagrancy/public intoxication (12.6%). 41.1% of subjects had no medical insurance. In the 12 months before last incarceration, 21.9% were hospitalized, 34.5% visited the ER, and 25.0% used EMS. However, during this period, only 27.1% visited a community mental health center (CMHC); 59.3% were seen by a psychiatrist (mean visit=5.3) and 34.0% by a social/case worker (mean visit=8.1). After release from jail, 24.4% of subjects visited a CMHC, 9.7% were hospitalized, and 15.7% used EMS.

Conclusion: Preliminary results showed high use of hospitalization and ER visits but low planned use of medical services among these subjects. Understanding patterns of resource utilization in schizophrenia may be useful in providing early and appropriate intervention to reduce recidivism in these individuals and avoid costly medical services.

Disclosures: LM, CB, RF, JF, and LA are employees of Janssen and are J&J stockholders.

Supported by funding from Janssen Scientific Affairs, LLC

14. Comparative Validation of the ISST-Plus, the S-STs, and the C-SSRS for Assessing Suicidal Thinking and Behavior

L. Alphs¹, D.V. Sheehan², L. Mao³, D. Williamson¹, E. Hammond⁴, R. May⁴, X. Li⁴, C. McCullumsmith⁴

¹Janssen Scientific Affairs, LLC, Titusville, NJ, USA, ²University of South Florida College of Medicine, Tampa, FL, USA, ³Janssen Research and Development, LLC, Titusville, NJ, USA, ⁴University of Alabama at Birmingham, Birmingham, AL, USA

Methodological Question Being Addressed: Convergent psychometric validation of the ISST-Plus, Sheehan Suicide Tracking Scale and C-SSRS

Introduction: Currently, the Columbia-Suicide Severity Rating Scale (C-SSRS) is designated by FDA as the standard for the assessment of suicidal ideation and behavior (SIB) for US regulatory data collection. Nevertheless, the phenomenology of SIB is complex and alternative approaches to collecting information that address needs that cannot be fully addressed with the C-SSRS are valuable. The InterSePT Scale for Suicide Thinking-Plus (ISST-Plus) and the Sheehan Suicide Tracking Scale (S-STs) have demonstrated validity as 2 alternative scales for collecting related SIB information. If the latter scales are used in trial for US regulatory submissions, it is important that they be mappable to FDA-mandated categories so that cross-study data summarization and comparison can be made. This study examines the concurrent validity for mapping these 3 scales to the FDA categories proposed in September 2010 and in August 2012.

Methods: Experienced raters from an academic clinical setting were formally trained in the use of the ISST-Plus, the S-STs, and the C-SSRS, and interrater reliability was confirmed during training. Subjects with SIB of varying levels of severity (“not at all suicidal” to “extremely suicidal”) were sampled from inpatient and outpatient settings and were consented to receive 3 separate SIB interviews by independent raters to gather information. Interview sequences for the 3 scales were randomized for each subject. Subjects were videotaped if separate consent was obtained. At the conclusion of each interview the appropriate scale was mapped to categories outlined in the FDA

draft guidance documents. Symptom mapping consistency of the 3 scales to the SIB categories is reported for each category.

Results: Five raters completed interviews of 45 subjects. Consistency of ratings for each category was identified for each subject on the ISST-Plus, the S-STIS, and the C-SSRS. Nine categories were mapped, including suicide attempt; aborted attempt; interrupted attempt; preparatory acts toward imminent suicide behavior; overall suicidal ideation; ideation: passive; ideation: active (nonspecific—no method, intent, or plan); ideation: active (method, but no intent or plan); ideation: active (method, intent, but no plan); ideation: active (method, intent, and plan); self-injurious behavior, intent unknown; not enough information (fatal); self-injurious behavior, no suicide intent; other (accidental, psychiatric, medical), no deliberate self-harm; not enough information (nonfatal).

Conclusion: The ISST-Plus, the S-STIS and the C-SSRS each represent potentially valuable instruments for identifying and categorizing SIB. Differences in categorization observed for the 3 scales may be related to variability in patient reports, rater reliability or differences in the instruments' approach to collecting SIB data. Additional psychometric work including interrater and intrarater reliability is necessary to support the validation of each of these scales.

Disclosures: LA, DW, and LM are employees of Janssen and J&J stockholders. EH, RM, XI, and CM are employees of the University of Alabama at Birmingham. DVS is a consultant who has received grant/research support and provided lectures/presentations for Janssen Pharmaceutica. A comprehensive list will be provided for the poster if accepted.

Support: Janssen Scientific Affairs LLC

15. Development of a New Patient-Reported Outcome (Pro) Measure for Major Depressive Disorder: Results of a Consortium-Based Approach

Blum, SI¹; Greco, N²; Martin, ML³; Hayes, RP⁴; Coons, SJ⁵

¹ Forest Research Institute, Jersey City, NJ; ²Abbott Laboratories, Abbott Park, IL; ³Health Research Associates, Mountlake Terrace, WA; ⁴Eli Lilly and Company, Indianapolis, IN; ⁵Critical Path Institute, Tucson, AZ

Methodological Question Being Addressed: Development and qualification of a new patient-reported outcome (PRO) measure for use in major depressive disorder (MDD) clinical trials.

Introduction: The U.S. Food and Drug Administration (FDA) has issued two guidance documents pertaining to development and qualification of PRO measures intended for use in medical product development to support labeling claims. The first guidance discusses the criteria by which FDA evaluates PRO instruments and their use.(1) The second (draft) guidance describes the FDA's process for qualification of drug development tools (DDTs), including PRO measures.(2) Presently we describe the process and progress to date of the Depression Working Group within the Critical Path Institute's (C-Path) PRO Consortium; a multi-stakeholder approach to develop and qualify a new MDD symptom-based PRO instrument.

Methods: The initial stage of the instrument development process incorporates several work streams: a systematic review of existing MDD instruments; a literature review of published studies describing patient experience with MDD; and input from an advisory panel of clinical/methodological experts. These work streams resulted in the development of a protocol and interview guide for in-depth interviews designed to elicit those concepts most important to MDD patients and the language used by patients to describe their symptoms. Following qualitative analysis of the interview transcripts, a consensus-building meeting was held to identify the relevant concepts to be measured, evaluate existing PRO measures and subsequently to guide the development of a draft item-pool suitable for further evaluation using both qualitative (cognitive interviews) and quantitative techniques.

Results: The instrument review identified 138 articles, from which 42 articles describing the development and/or testing of MDD symptom measures were retained. These articles addressed 26 existing instruments, from which 13 were evaluated in-depth to determine the concepts measured and the role of patients in the development process. The literature review resulted in identification of 177 articles, from which 19 were reviewed that identified emotional, physical and cognitive symptoms of depression, disease-related impacts, and other signs/concepts for consideration by the advisory panel and for use in developing the concept elicitation interview guide. A total of 40 in-depth one-on-one interviews were conducted with subjects with MDD (mean age [SD]: 46.2 [11.8], 67.5% Female) from a broad representation of educational, socio-economic and ethnic/racial backgrounds. Saturation was achieved with the first 32 coded transcripts. Following the decision to develop a new measure, a preliminary draft

measure containing 36 items was created by the research team. Additional qualitative and quantitative research will further test and refine this draft instrument prior to submission for qualification by FDA.

Conclusions: A consortium-based approach has been able to successfully develop a new draft PRO measure for MDD, which incorporates evidence from published literature and insights from qualitative interviews to reflect the patient's voice and perspective. This approach is likely to satisfy best research practices and current FDA guidance for PRO instrument development and qualification.

Disclosures: Funding for this research was provided by the following PRO Consortium member firms: Abbott Laboratories; Bristol-Myers Squibb; Eli Lilly and Company; Forest Laboratories; Janssen; Pfizer; Shire, and Sunovion Pharmaceuticals.

References: (1) Guidance for Industry - Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Final version, December 2009

(2) Guidance for Industry - Qualification Process for Drug Development Tools, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Draft version, October 2010

16. An Examination into the Effects of Five Rater Training Modalities on the Project Conduct in International AD Trials

Erin B Kornsey¹, Bethanne Friedmann¹, Edward Bartolic¹, Sharon Carter¹, Cordelia Zakrajsek¹, Neal R Cutler², Henry J Riordan¹

¹ *WORLDWIDE CLINICAL TRIALS, King of Prussia, PA*, ² *WORLDWIDE CLINICAL TRIALS, Beverly Hills, CA*

Methodological Question Being Addressed: Which rater training modality is the most effective in capturing accurate data and reducing project team effort?

Introduction: The aim of this study is to investigate the effectiveness of different rater training modalities data quality and study conduct for two early phase Alzheimer's disease clinical drug trials.

Methods: In order to determine the relative superiority of selective training methods on data accuracy operationalized in a blinded review of deviation, violation and query rates various rater training methodologies were evaluated across two early phase double blind, randomized, placebo controlled clinical trials that were designed to assess the effects of a novel drug on Alzheimer's disease. Specifically five rater training modalities were examined across the two studies: training at the investigator meeting only (IM) (n=35), self paced web portal training (SW) which entails raters reviewing didactic presentations, videos and then taking subsequent quizzes on a special project website (n=46), onsite face to face training provided by an expert rater (OS) (n=10), group web based conference training where a number of sites attended a live presentation simultaneously (GW) (n=16), and individualized web based conference training which entails an individual rater or site attending a live presentation given by an expert rater (IW) (n=34). Rater errors were examined by clinician's review of the data as well as electronic data capture (EDC) database queries.

Results: From a total number of 36 sites, data from 143 raters across 5 countries were evaluated and included in the analyses. Sites whose raters were trained by an onsite visit by the expert rater yielded no subsequent data queries regarding the clinical assessments with no deviations, violations or queries regarding the clinical assessments by the clinician review.

Conclusions: Onsite rater training provided by an expert rater appears to be to be the most effective method of ensuring data is captured correctly in a timely fashion using query rates and severity or those rates as the dependent measure. Raters who completed training via self paced web portal didactics and tests exhibited the highest number of data queries (e.g. possible unblinding of rater, incomplete certification procedures by sites, etc.) Raters who attended the web based group conference training as a group or individual had a relatively larger number of queries of varying severity. It is not yet known if the differences in query rates and severity evidenced among the various rating training methodologies were indicative of effect sizes differences as a measure of the individual site's ability to successfully differentiate drug from placebo effects and site data on this will be examined once all data is unblinded.

Disclosures: The authors report no conflicts of interest for this work.

17. Evidence That Sites Can Conduct High Quality Interviews and Ratings in Global Clinical Trials

David G. Daniel, MD¹, Alan Kott, MD¹, Amir Kalali, MD²

¹Bracket Global, LLC, ²Quintiles

Methodological Question Being Addressed: Can sites provide high quality subject interviews and ratings procedures in global clinical trials?

Introduction: As drug-placebo differences have diminished over time in global CNS clinical trials, the quality of interview and rating procedures performed by sites have come under increasing scrutiny (1-4). We describe the pooled results of surveillance measures of interview and ratings quality in ten separate international clinical trials.

Method: A proprietary video/audio recording system was utilized to record PANSS rating procedures in ten schizophrenia clinical trials conducted in North America, Europe, South America and Asia. Sites uploaded video or audio recorded ratings assessments for review by calibrated external reviewers. External reviewers provided feedback on ratings quality on an ongoing basis to the site and sponsor. Prior to study initiation, raters were trained at investigators meetings by highly interactive procedures, including slide presentations, rating of videotaped patient interviews, and, in some cases, interview and rating of live actors trained to portray schizophrenia symptoms. Instruments used to assess ratings and interview quality included the Rater Quality Questionnaire (RQQ) (5) and the Research Interview Assessment Scale (RISA) (6).

The RQQ consists of two items addressing: 1) the overall quality of the patient and/or informant material collected; and 2) the overall quality of the ratings or diagnostic evaluation (including proper application of the rules and anchor points of the rating scale or structured interview) for the current subject visit. Each domain is evaluated on a Likert-like scale (1-3). A score of 1 represents no deficiencies in information gathering and/or scoring. A score of 2 represents deficiencies that have only a minimal impact on the scoring. A score of 3 reflects serious deficiencies in either collecting the information or rating the symptoms observed. The RISA is a 16 item scale representing 4 domains of interview quality. Higher scores represent better quality interviews.

Results: For the RQQ analysis, the initial pooled dataset consists of 1187 evaluated interviews. The mean RQQ scores for sufficiency of the data collected by interview and competency in use of the rating scales were $1.35 \pm .51$ and $1.31 \pm .50$, respectively. 97.5% of the interviews were graded as adequate with no more than minimal deficiencies in any area. Only 1.4 % of interviews had serious deficiencies in one of the domains and only 1.1% of interviews had serious deficiencies in both domains. Most (60.2%) interviews were considered as fully proficient in both information gathering and adherence to PANSS scale instructions. The remainder had at least minor deficiencies. 7.2% of interviews had good quality of information gathering but had minor difficulties in scoring this information; 10.8% of interview had minor deficiencies in information gathering but the information was scored using proper technique. 19.5% of interviews had minor deficiencies in both information gathering and scoring.

1442 videotaped PANSS administrations at sites were graded for interview quality by external reviewers using the RISA. 64.6% were regarded as excellent (RISA Score 28-30), 28.9 % as acceptable (RISA score 24-27) and only 6.5% as poor or unacceptable.

Discussion: The initial data indicate that in global clinical trial settings with surveillance of ratings and ongoing feedback to investigators the quality of interview data and proficiency of ratings were judged to be adequate or better by external reviewers in the large majority of cases. These results are preliminary and additional data will be reported from ongoing and recently completed studies. This study's findings differed from those of Jeglic et al (2007) and Engelhardt et al (2006) in which the majority of interviews were judged to be deficient. We speculate that rapid feedback to the raters from the external reviewers may have been useful in maintaining interview and ratings quality in the current studies.

References: 1) Kemp AS, Schooler NR, Kalali AH et al: What is Causing the Reduced Drug-Placebo Difference in Recent Schizophrenia Clinical Trials and What Can be Done About It? *Schizophrenia Bulletin* (2010) 36 (3): 504-509.

2) Engelhardt N, Feiger AD, Cogger KO, et al: Rating the raters: assessing the quality of Hamilton rating scale for depression clinical interviews in two industry-sponsored clinical drug trials. *J Clin Psychopharmacol* (2006) 26 (1):

71–74.

3) Jeglic E, Kobak, KA, Engelhardt N et al: A novel approach to rater training and certification in multinational trials. *Int Clin Psychopharmacol* (2007) 22 (4): 187–191.

4) Kobak KA: Inaccuracy in Clinical Trials: effects and methods to control inadequacy. *Curr Alz Res* 2010 Nov; 7 (7) :637-641

5) Daniel DG and Kott A: Initial Findings of the The Rater Quality Questionnaire (RQQ): A New Tool for Evaluating Quality of Ratings in Global Clinical Trials. Poster to be presented at the 25th ECNP Congress, October 13-17, 2012, Vienna, Austria.

6) Spear C, McNamara C, Miller D et al: A New Scale for Evaluating Interviewing Competency. Proceedings of the New Clinical Drug Evaluation Unit Annual Meeting, Boca Raton, Florida, June 14, 2006.

Funding: Statistical analyses were funded by Bracket Global, LLC.

18. The Impact of Audio/Video In-study Interview Monitoring Implementation on Subject Recruitment

Alan Kott, MD, David G. Daniel, MD

Bracket Global, LLC

Methodological Question Being Addressed: Does implementation of audio/video site monitoring methodology in a clinical trial impact recruitment of subjects?

Introduction: In the perspective of diminishing drug-placebo differences in current clinical trials in schizophrenia sponsors are implementing various monitoring methodologies including audio/video recordings of individual subject interviews to increase the scrutiny of individual ratings. Research sites are often reluctant to utilize these methodologies mainly because of the fear that having such a methodology in place could affect recruitment. We present a pooled analysis of recruitment at individual research sites in eight separate international clinical trials to address the question whether by implementing audio/video monitoring methodology the recruitment at the sites is affected.

Methods: The proprietary audio/video recording system is being implemented in eight schizophrenia clinical trials. Sites are requested to record each screening/randomization PANSS subject interview and submit for evaluation to an independent reviewer. While the use of the system is not mandatory for the individual subjects, sites are strongly encouraged to utilize it for every subject screened and/or randomized depending on trial methodology. For the purposes of the analysis sites are considered compliant if they utilized the system for at least one subject screened and/or randomized. Sites that have not used the system for any of the subjects are considered non-compliant. Two tailed t-test analysis on all pooled data was performed to compare the number of subjects randomized at research sites by compliance.

Results: The overall data pool consisted of 2340 randomized subjects. Sites compliant with the audio/video monitoring system randomized on average 5.81 subjects compared to non-compliant sites with an average randomization of 3.52 subjects ($t=-6.41$, $df=508$, $p < 0.001$). Compliant sites randomized on average significantly more subjects even when we only analyzed sites with more than 5 subjects randomized.

Discussion: The initial analysis indicates that implementing audio/video monitoring system should not represent a significant obstacle in recruitment. The majority of sites compliant with the monitoring system randomized significantly more subjects compared to non-compliant sites. We believe that this finding reflects the difference between sites' investment in the clinical trial. Invested high recruiting sites who are aware of current issues in clinical trials are more likely to comply with the monitoring compared to the sites who are less invested and on average recruit significantly less subjects.

References: Alphs, Larry; Benedetti, Fabrizio; Fleischhacker, W. Wolfgang; Kane, John M. (2012): Placebo-related effects in clinical trials in schizophrenia: what is driving this phenomenon and what can be done to minimize it? In *Int J Neuropsychopharmacol*, pp. 1–12.

Daniel, David; Kott, Alan (2012): Initial findings of a method for monitoring and maintaining reliable and accurate ratings in global schizophrenia clinical trials. Poster presented at the American Psychiatric Association Annual Meeting, Philadelphia, PA, May 5-9, 2012.

19. Identification of an Integrated Negative Factor in Schizophrenia Using Item Response Theory

Khan A^{1,4}, Lindenmayer JP^{3,4}, Opler M^{1,3}, Yavorsky C^{1,2}

¹ProPhase, LLC, ²CROnos CCS, ³New York University School of Medicine, ⁴Nathan S. Kline Institute for Psychiatric Research, ⁵Pratt Institute

Methodological Question Being Addressed: Negative symptoms of schizophrenia remain an unmet therapeutic need. This presentation takes the ideas that surfaced from the NIMH consensus (2005) along with various factor structures of the Positive and Negative Syndrome Scale (PANSS) and assess these findings using Item Response Theory (IRT) to determine the best classification system for negative symptoms. The PANSS Negative Subscale and Negative Factor structures are the most commonly used in clinical trials assessing negative symptom endpoints. In developing and assessing medications that treat negative symptoms, this presentation asks if there is scientific evidence utilizing IRT to identify a robust, sound dimension for Negative Symptoms?

Background: Negative symptoms are a core presentation of schizophrenia and have significant predictive value for patients' community functioning. While their importance has been recognized, there has been a debate as to which aspects of psychopathology should be considered part of the negative symptom construct.

Objectives: (1) To examine which symptoms have been identified as negative symptoms through examination of various factor analyses, and (2) to examine the quality of these negative items using IRT.

Methods: PANSS scores from 7,187 schizophrenia patients who were assessed for enrollment in psychopharmacology trials were obtained. 1) First, a review of 24 published factor structures on the PANSS was conducted. All items loading on Negative Factors were identified. 2) Secondly, a Principal Components (PCA) with no rotation to assess dimensionality of the selected items was conducted on the dataset (n = 7,187). 3) A non-parametric IRT was conducted on the 7,187 patients (total PANSS score covered a broad spectrum of psychopathology from 30 - 161). Option characteristic curves (OCCs) and Item Characteristic Curves (ICCs) examined the probability of rating each option as a function of this Integrated Negative Factor.

Results: 1) Of the 24 published studies, 8 items (Emotional Withdrawal, Passive/Apathetic Social Withdrawal, Lack of Spontaneity, Poor Rapport, Blunted Affect, Motor Retardation, Active Social Avoidance, Disturbance of Volition) were identified as loading on $\geq 25\%$ (of the 24 studies). 2) Using the dataset (n= 7,187), PCA revealed one component with an eigenvalue ≥ 1 . The 'Negative Factor' shows the first eigenvalue is 4.871 times larger than the second component, confirming unidimensionality. 3) All items forming the Integrated Negative Factor performed very well and discriminate better along symptom severity compared to the Negative Subscale and commonly used negative factors. The average item information function (IIF) for the Integrated Negative Factor ranges from 0.15 to 0.30, and is higher than the IIF of 0.10 to 0.17 for the Negative Subscale and other commonly used negative factor structures.

Conclusions: The findings provide evidence for a Negative Factor across a broad spectrum of psychopathology. Findings further allowed for independent formation of a specific negative dimension from a commonly used rating scale, thus eliminating the need for adding a new measure to a clinical trial, which can place burden on the patient and rater. This may inform the current debate about revised classification systems of negative symptoms.

Keywords: Schizophrenia, Negative Symptoms, PANSS

Disclosures: No authors have conflicts of interest related to this study to disclose. Khan A is employed part time at Manhattan Psychiatric Center and at ProPhase LLC, and volunteers at Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY. Yavorsky, C is a full time employees of CROnos CCS, Opler M is the owner of ProPhase LLC and employed at New York University.

20. Development of the Dynamic Social Cognition Battery (DSCB): A Comprehensive Toolkit Utilizing Dynamic Images to Assess Social Cognition in Schizophrenia and Related Disorders

Khan A^{1,4}, Rothman B¹, Lucic L^{1,5}, White J¹, Yavorsky C^{1,2}, Opler M^{1,3}, DiClemente G², Jovic S¹

¹ProPhase, LLC, ²CROnos CCS, ³New York University School of Medicine, ⁴Nathan S. Kline Institute for Psychiatric Research, ⁵Pratt Institute

Methodological Question Being Addressed: Social cognition has been broadly studied in autism spectrum, and includes impairments in: (a) affect perception; (b) social perception; (c) attributional style; and (d) theory of mind (ToM). Deficits in social cognition contribute directly to social dysfunction leading to impaired communication and difficulties in community functioning. Social cognition as a target for future pharmacotherapy has the potential to provide meaningful functional improvements in outcome in schizophrenia (Sz) and affective disorders, and has become a high priority for developing future psychosocial and pharmacological treatments. Can a comprehensive kit using dynamic images be a better predictor and more sensitive to change of social cognition in research trials?

Background: The complexity of the constructs of social cognition makes it difficult to develop adequate assessment tools and results are needed to address the disorder specificity of these impairments. Studies examining processes of social cognition have traditionally relied upon the use of static (photographic or written) stimuli in order to assess facial displays of emotion and ToM which are dynamic by nature. Ecological validity of such stimuli is debatable. There is a growing evidence of specialized brain systems that are preferentially activated by “biological motion” stimuli (including moving faces) giving further support to the need for dynamic stimuli.

Objectives: (1) To describe the development of a comprehensive toolkit to assess social cognition utilizing dynamic images for Emotion Perception, ToM and Attributional Styles.

Methods: The authors present the developmental stages that captures social cognition on three levels. To prepare this measure, empirical research were examined to review: (1) the functional approach to Social Cognition, and (2) preliminary findings that serves as a foundation for developing the current measure.

Results: Using empirical evidence and in collaboration with an interdisciplinary team, the DSCB consists of: 1. Facial and Auditory Emotion Perception Scale (FAEPS) measures i) Emotion Identification, ii) Emotion Discrimination, iii) Verbal and Non-verbal Emotion Processing. Ratings of the 5 basic emotions: happy, surprise, anger, disgust, sad. Part 1: 20 faces as short video clips. Part 2: 10 non-verbal short videos clips. Part 3:10 verbal short story clips assessing auditory emotion processing. 2. Dynamic Images ToM Task (DI-ToM) contains video clips corresponding with the 3 main ToM stages: (a) precursors of ToM (e.g., emotion recognition), (b) first manifestations of a real ToM (e.g., understanding of false belief), and (c) mature aspects of ToM (e.g., second-order beliefs). DI-ToM matches in linguistic demands and assessed Diverse Desires, Diverse Beliefs, Knowledge Access, False Beliefs, Hidden Emotions. 3. Test of Attributional Style Task (TAS) is divided into 2 parts: Part 1: 16 scenarios (8 positive, 8 negative) through short video clips. Additionally, error profiles from the FAEPS will also be computed to assess attributional style.

Conclusions: This battery is designed to serve as an outcome measure for clinical trials of social cognitive enhancing drugs, social cognition training, and as a reference point. This battery makes it possible for researchers worldwide to accurately assess a variety of complex social cognitive processes in related psychiatric disorders via software.

Keywords: Social Cognition, Neurocognition, Scale Development

Disclosures: No authors have conflicts of interest related to this study to disclose. Yavorsky, C and DiClemente G are full time employees of CRonos CCS, Opler M is the owner of ProPhase LLC; Khan A is employed part time at Manhattan Psychiatric Center and at ProPhase LLC, and volunteers at Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY. Rothman B, White J and Jovic J are full time employees at ProPhase LLC. Lucic L is a full time employee at ProPhase LLC and a visiting professor at Pratt Institute. NY, NY.

21. A Cognitive Task Sensitive to Dentate Gyrus Activity which has Implications for Assessing Neurogenesis Status in Various Conditions Including Normal and Pathological Ageing

Keith Wesnes

Bracket Global, Goring, UK & Swinburne University, Melbourne, Australia

Methodological Question Being Addressed: This poster addresses the issue as to whether a cognitive test can act as a biomarker of dentate gyrus activity, and can thus be reflective of hippocampal neurogenesis activity.

Introduction: The seminal discovery that the human dentate gyrus (DG) retains its ability to generate neurons throughout life has raised the possibility that therapies could be developed to protect or promote this neurogenesis as it deteriorates due to ageing, insult and disease. The DG plays a crucial role in associative memory, and the

degenerative changes which compromise neurogenesis in the DG are believed to contribute to memory disturbances in normal ageing and the early stages of AD. Performance on pattern separation tasks has been demonstrated to be under the control of the DG, and fMRI studies have identified the DG to be highly and selectively active when volunteers perform difficult visual object pattern separations. The CDR System picture recognition task assesses visual object pattern separation, and in a cohort of over 3000 volunteers aged 18 to 87 years, a selective, marked and highly significant age-related decline was identified in the ability to discriminate originally presented pictures from different but very similar pictures. Further, patients with Mild Cognitive Impairment were shown to be selectively inferior on this DG sensitive discrimination compared to age-matched healthy controls. This task thus offers the opportunity to assess DG activity in clinical trials, and thus potentially reflect neurogenesis activity. The aims of the present study were to confirm and extend these findings by investigating over 47,000 individuals tested over the internet, as well as to determine whether this task would provide evidence of compromised neurogenesis in clinical populations including chronic pain, Parkinson's disease and depression.

Methods: Data from 47,731 individuals aged 5 to 102 years who performed the CDR System picture recognition task over the internet were analysed. Data from this task in clinical populations including chronic pain, Parkinson's disease late life depression were also evaluated.

Results: This study confirmed the original pattern with regard to the decreased ability with adult ageing to discriminate the originally presented pictures from the very similar ones; and extended our knowledge by revealing that younger children were also compromised in this ability. Further, the declining ability to discriminate the pictures with ageing was also associated with selectively longer reaction times, extending our understanding of the phenomenon of declining pattern separation accuracy by showing that it is accompanied by declining speed. Finally, other analyses revealed that patients with chronic pain, Parkinson's disease and depression were also selectively impaired on this task.

Conclusions: Hippocampal neurogenesis declines with normal ageing, and the results of the present study are consistent with this; also suggesting that it develops from 5 to 15 years. Further both animal and autopsy studies indicate the neurogenesis in the dentate gyrus is compromised in chronic pain, Parkinson's disease and depression, and again, the findings from the picture recognition task support this. The task thus provides an opportunity to monitor DG activity in various clinical populations, and could be a useful tool in evaluating compounds aimed at promoting, maintaining or restoring neurogenesis. The opportunity to study large populations via the internet has applications to the various long-term patient registries being set up to study preclinical dementia.

Disclosure: The author is employed by a company which provides the test as a service to the clinical trial industry.

22. An Automated Executive Function Task for Repeated Administration in Clinical Trials in Schizophrenia

Keith Wesnes^{1, 2}, Richard Wojciak¹, Todd Feaster³, Cynthia McNamara³, Jean Dries³, Adam Butler³, David Krefetz⁴, Diana Pittman⁴, Jessica Fogel⁴, Shirley Kramer⁴, Erica Ridolfi⁴, Lawrence Brownstein⁴, Howard Hassman⁴

¹Bracket Global, Goring on Thames, UK, ²Swinburne University, Melbourne, Australia, ³Bracket Global, Wayne, USA, ⁴CRI Lifetree, Mount Laurel, NJ, USA

Methodological Question Being Addressed: This poster addresses an important issue in clinical research, the development a test of executive function for use in schizophrenia trials which is not subject to training effects, and is also clinically relevant.

Introduction: Executive function, also known as executive control, is a set of cognitive abilities that control and regulate other aspects of function to facilitate goal-directed behaviour. A number of cognitive tests capture essential aspects of executive function including the Trail Making and Wisconsin Card Sorting tests; but can be subject to notable training effects when administered repeatedly which limit their utility in clinical trials. We have adapted and automated a rule switching task of executive function which has previously been shown to involve activation of the medial and dorsolateral areas of the frontal cortex; and have validated it for use in volunteers. In the present study we studied its utility, validity, sensitivity and suitability for repeated use in patients with schizophrenia.

Methods: Thirty male and female patients with schizophrenia who were on stable medication were each tested on three separate days. On the first day the severity of disease was assessed using the CGI-S and the UPSA-B was administered. On this and the next two study days the patients performed the CDR System, the CDR System

Executive Function test (EFT), the Trail Making Task (TMT) and the NAB Mazes test. Performance on the task was compared to that of healthy volunteers. Training effects were evaluated using mixed model repeated measures analysis, and correlation procedures were used to examine test-retest reliability and the relationship to the UPSA-B.

Results: Significant and large effect sized improvements due to repeated administration were seen on NAB Mazes and all measures from the TMT; but not the CDR System nor the CDR EFT. The CDR EFT correlated appropriately with the TMT and NAB Mazes. Test-retest scores were actually higher for the TMT and NAB Mazes over the study sessions than for the CDR EFT, indicating that test-retest reliability does not predict stability with repeated testing. The CDR EFT and TMT both showed large effect sized impairments compared to healthy volunteers. The Coefficient of Variance of the CDR EFT was over 30% smaller than for the TMT or NAB Mazes, which suggests that the CDR EFT will be more sensitive to treatment effects. The UPSA-B total score correlated significantly with CDR EFT ($r=0.57$) but not with NAB Mazes or any of the TMT measures.

Conclusions: Overall, the CDR EFT shows promising statistical properties for the essential purpose of repeated administration in clinical trials. The test identifies large effect sized deficits in patients with schizophrenia and also has clinical and everyday relevance; evidenced by correlating firmly with the UPSA-B. This new executive function tests should thus prove a valuable instrument for clinical trials of cognition enhancers in patients with schizophrenia.

Disclosures: All authors are employed by companies which provide clinical trial services to the pharmaceutical industry

23. Assessment of Cognition in Smoking Abstinence and Nicotine Replacement Studies: Review of Methodological Issues and Study Results in Healthy Controls and in Chronic Schizophrenia

Kingery, LR, Jaeger, J, Maruff, P

CogState.

Methodological Question Being Addressed: Discussion and review of best methods and design issues in the assessment of cognition in smoking cessation studies, and comparison of results of 2 studies, one in healthy controls and one in chronic schizophrenia.

Introduction: The purpose of this research is to review the methodological issues involved in assessment of cognition in smoking cessation studies, including previous findings, design considerations, and future directions, presented in the context of the results of two nicotine abstinence studies, one in healthy controls and one in chronic schizophrenia.

Methods: Review of literature on cognition assessment and presentation of results of 2 studies. The first study included $n=90$ healthy volunteers enrolled in a cross-over design with a 2-week washout. Assessments were conducted at Baseline ~12 hours day 1 of smoking abstinence, Day 2 pre-dose, and 3, 6, 9, and 12 hours post nicotine replacement dose. In the study of chronic schizophrenia, the design was parallel group, randomized, placebo controlled with assessments at pre-dose and at 1, 2, and 3 hours post dose nicotine replacement.

Results: Results are discussed in the context of previous findings regarding cognitive domains impacted by smoking abstinence and nicotine replacement. Differential impact on cognitive functioning was observed in both the healthy volunteers and the chronic schizophrenia groups, particularly on measures of reaction time and information processing speed. Results are presented in terms of standardized effect sizes to facilitate comparison across similar studies. Findings regarding dosing and time-course of effects will be presented.

Conclusions: Assessment of the cognitive effects of smoking abstinence and nicotine replacement treatments requires careful attention to several methodological issues and are informed by known effects on different cognitive domains. Implications of the effects of smoking abstinence and nicotine replacement on relevant nootropic clinical development is discussed.

Disclosures: All authors are employees of CogState.

24. Improved Statistical Power of Alzheimer Clinical Trials by Item Response Theory: Proof of Concept by Application to the Activities of Daily Living Scale

Steven D. Edland

Department of Neurosciences, University of California San Diego, La Jolla, CA.

Methodological Question Being Addressed: "Can simple adjustments to outcome measures improve the statistical efficiency of clinical trials?"

Introduction: The Alzheimer's Disease Cooperative Study-Activities of Daily Living inventory (ADCS-ADL) (ADAD. 1997; 11:S33-39) is a frequently used functional endpoint in clinical trials for AD that assesses patient functional ability on the basis of informant ratings of patient performance on a variety of everyday tasks. Previous research has shown that the items comprising the ADCS-ADL are sensitive to characteristic longitudinal trajectories in AD. However, standard procedures for combining information from individual items into an overall test score may not make fully efficient use of the information provided by informant responses.

Methods: The current study explored an application of item-response theory (IRT) techniques to the calculation of test scores on the ADCS-ADL using data from two ADCS clinical trials of mild to moderate AD patients.

Results: We found that IRT-based scoring increased sensitivity to change in functional ability and improved prospective statistical power of the ADCS-ADL as an outcome measure in clinical trials. A trial using IRT rescored ADCS-ADL items would require 15% less subjects than a trial using standard scoring.

Conclusions: This proof of concept investigation suggests that substantial improvements in efficiency may be possible by IRT methods. Development of methods to further optimize IRT rescoring of clinical trial endpoints is ongoing.

25. An Open-Label Randomized Exploratory Investigation of Risperdal* Consta* and Oral Antipsychotic Therapy in the Treatment of Early Psychosis

Quinn AM¹, Camacho, F², Mitchell, D¹, Chue, P³, Malla, A⁴

¹Janssen Inc., Toronto, ON, Canada, ²Damos Inc., Toronto, ON, Canada, ³University of Alberta, ⁴McGill University

Methodological Question Being Addressed: Exploratory investigation of the use of long-acting vs. oral antipsychotic therapy in early psychosis.

Introduction: Long-acting injectables offer distinct benefits vs. oral therapy in the management of schizophrenia¹, but data supporting their use in earlier disease is more limited. This pilot study explored the impact of oral second-generation antipsychotics and long-acting risperidone injections in subjects with recent onset psychosis to inform further confirmatory investigations.

Methods: This was a 24 month, open-label, 2-arm randomized hypothesis-generating study. Eighty-five subjects with early onset (≤ 3 years) DSM-IV diagnosis of schizophrenia, schizophreniform disorder or schizoaffective disorder were randomized to either risperidone long-acting injections (RLAI) every 2 weeks or continuation of therapy with the subject's current oral atypical antipsychotic (risperidone, olanzapine or quetiapine). The study began with an 18 week stabilization phase which was followed by an 82 week maintenance phase for both arms. Primary evaluation was based on the Positive and Negative Syndrome Scale (PANSS), time to relapse as defined by Csernansky, and the Social and Function Outcomes (SOFAs). The safety and tolerability of Risperdal* Consta* was also explored. This was a pilot study; no formal sample size calculations were performed.

Results: One hundred and one (n=101) patients were screened at 12 Canadian centers; 85 of these subjects were randomized. The primary analysis included 77 patients who received at least three doses of risperidone long-acting injections (n = 42; 54.5%) or six weeks of oral antipsychotic therapy (n = 35; 45.5 %), in addition to having at least one post-baseline efficacy assessment.

The number of weeks to stabilization was 11.7 in both the risperidone LAI and oral antipsychotic groups, ($p=0.7870$). Of the 63 patients who entered the maintenance phase of the study, 34.4% (11/32) risperidone LAI and 16.1% (5/31) oral antipsychotic patients relapsed, although the study was not powered to show a difference between these values. It was not possible to calculate the median time to relapse. PANSS scores improved in each group. Similar proportions of stabilized risperidone LAI versus oral antipsychotic patients did not complete the study 17/32 (53.1%) versus 16/31(51.6%), respectively. However, non-compliance contributed to 17.6% (3/17) of the discontinuations in stabilized risperidone LAI patients versus 50.0% (8/16) of the discontinuations in stabilized

oral antipsychotic patients. Safety was assessed in the 85 randomized patients, and no new or additional signals were observed.

Conclusion: This 24-month Canadian study explored the clinical outcomes, safety, tolerability, and resource utilization between risperidone LAI and oral antipsychotics in recent onset psychosis to inform future confirmatory trials. Further investigation of long-acting and oral therapy in early and first-episode psychosis is recommended.

Supported by Janssen Inc.

References: 1. Gaebel W, Schreiner A, Bergmans P, et al. Relapse Prevention in Schizophrenia and Schizoaffective Disorder with Risperidone Long-Acting Injectable vs Quetiapine: Results of a Long-Term, Open-Label, Randomized Clinical Trial. *Neuropsychopharmacology*. 2010;35(12):2367–2377.

26. Duration of Marijuana Use and the Cognitive Impact on Simple Reaction Time

Harrington MN¹, MorganDM¹, Goddard SG², Gray KM³

¹CNS Vital Signs, LLC, ²Texas A&M University Department of Statistics, ³Medical University of South Carolina

Introduction: Research on the acute, residual and long-term cognitive effects of cannabis has yielded mixed findings (1,2,3). This study sought to determine if there was a relationship between cannabinoid urine level, age and duration of marijuana use with simple reaction time as measured by a computerized neurocognitive test that captures reaction time with millisecond (mS) precision.

Methods: At the baseline visit before treatment randomization, in the parent marijuana cessation trial, 98 participants completed a comprehensive computerized neurocognitive testing battery (CNS Vital Signs)(4). The group was split into four 2 year cohorts, Group A- those admitting to smoking less than 2 years, Group B- 2 or more years but less than 4 years, Group C- 4 or more years but less than 6 years, and Group D- greater than 6 years. Only one subject reported smoking 8 or more years. The sample ranged in age from 15 to 21 years old.

Simple reaction time was captured using a computerized version of the Stroop Test. The Stroop Test, is a well-known executive function task. In capturing simple reaction time, participants were asked to press the space bar as soon as they see any word stimulus on the computer screen. There are no correct or incorrect responses, only “simple” responses to presentation of a word stimulus. A urine creatinine-normalized cannabinoid test was used to reliably compare quantitative levels (urine cannabinoid level/urine creatinine level). A regression analysis was utilized to test for different mean reaction times between the four duration-of-use categories, including age and urine cannabinoid levels as covariates. Invalid tests for both the urine cannabinoid test and the Stroop Test were removed prior to analysis.

Results: There was a significant difference in the “years smoked” groups, where simple reaction time was the dependent variable. Means and standard deviation for the years smoked groups were as follows. Group A 268 mS (30). Group B 255 mS (33). Group C 264 mS (47). Group D 334 mS (171).

Conclusions: Participants with the longest duration of marijuana use had slower simple reaction times. This finding suggests that reaction times may increase as years of smoking marijuana increases, even after age and cannabinoid urine levels are controlled for.

This research was supported by NIDA grant R01DA026777 and NCRR grant UL1RR029882.

27. Cognitive Remediation in Schizophrenia Trials: Changes on Individual Cognitive Tests and Domains Within the Control Group

Rajiv Radhakrishnan M.D.¹, Jack Tsai Ph.D.^{2,3}, Brian Kiluk Ph.D.²

¹Department of Psychiatry, Zucker Hillside Hospital, Glen Oaks, NY; ²Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA; ³VA New England Mental Illness Research, Education, and Clinical Center, West Haven, CT

Methodological Question Being Addressed: What is the degree of pre-post change seen in individual cognitive tests and in each cognitive domain within the control group of cognitive remediation training (CRT) trials of schizophrenia?

Introduction: Various meta-analyses on randomized controlled trials of CRT for schizophrenia have shown small-to-moderate treatment effects on cognitive test performance (Cohen’s $d=0.38-0.45$). However, there is wide

variability in research design and type of control group in CRT trials, and little examination of the change in cognitive test performance that occurs within control groups.

Aim: To examine pre-post change in cognitive test performance reported within control groups in CRT trials of schizophrenia; and to examine the change with individual cognitive tests and in different cognitive domains.

Methods: Of 42 randomized control trials of CRT for schizophrenia that met rigorous methodological criteria, 30 studies provided adequate data to calculate the effect-size of pre-post change (Cohen's d) on cognitive test performance within the control group. Studies which used the same cognitive test were grouped together to derive a mean effect-size of change for individual tests. The individual tests were categorized according to the cognitive domain that they tapped into and mean effect-size of change for different domains was calculated.

Results: The 30 studies comprised a total of 792 patients with a mean age of 35.34 ± 7.22 years and mean IQ of 92.11 ± 6.08 . The mean effect size of pre-post change (Cohen's d) in the control group was 0.15 ± 0.14 (range = $-0.07 - 0.52$; sample size weighted Cohen's $d = 0.15$; inverse-variance weighted Cohen's $d = 0.15$). For individual cognitive tests, the mean effect-size of change was smallest for Digit Span ($d = 0.045 \pm 0.036$). Degraded-CPT showed smaller change ($d = 0.053 \pm 0.064$) than simple CPT ($d = 0.081 \pm 0.132$). The cognitive test with largest effect-size of change was Backward Masking task ($d = 0.51 \pm 0.02$). The cognitive domain with the smallest mean pre-post effect-size of change was Speed of Processing ($d = 0.065 \pm 0.039$), while the domain with largest effect-size of change was Social Cognition ($d = 0.55 \pm 0.19$).

Conclusions: The control groups in CRT trials show small effect-size changes in cognitive test performance over time. Individual cognitive tests show different degrees of susceptibility to non-specific change as measured by pre-post change within the control group. The profile of improvement in different cognitive domains within the control group may represent the degree of improvement that can be expected due to non-specific change in studies of CRT in schizophrenia. The wide variability in test measures used across studies is a limitation of this analysis.